CONVERGENCE PROPERTIES OF THE BFGS ALGORITHM*

YU-HONG DAI†

Abstract. The BFGS method is one of the most famous quasi-Newton algorithms for unconstrained optimization. In 1984, Powell presented an example of a function of two variables that shows that the Polak–Ribière–Polyak (PRP) conjugate gradient method and the BFGS quasi-Newton method may cycle around eight nonstationary points if each line search picks a local minimum that provides a reduction in the objective function. In this paper, a new technique of choosing parameters is introduced, and an example with only six cyclic points is provided. It is also noted through the examples that the BFGS method with Wolfe line searches need not converge for nonconvex objective functions.

Key words. unconstrained optimization, conjugate gradient method, quasi-Newton method, Wolfe line search, nonconvex, global convergence

AMS subject classifications. 65K05, 65K10

PII. S1052-6234(01)38345-5

1. The BFGS algorithm. The BFGS algorithm is one of the most efficient quasi-Newton methods for unconstrained optimization:

\[ \min f(x), \quad x \in \mathbb{R}^n. \]

The algorithm was proposed by Broyden [2], Fletcher [5], Goldfarb [7], and Shanno [19] individually and can be stated as follows.

Algorithm 1.1. The BFGS algorithm.

Step 0. Given \( x_1 \in \mathbb{R}^n; B_1 \in \mathbb{R}^{n \times n} \) positive definite; Compute \( g_1 = \nabla f(x_1) \). If \( g_1 = 0 \), stop; otherwise, set \( k := 1 \).

Step 1. Set \( d_k = -B_k^{-1}g_k \).

Step 2. Carry out a line search along \( d_k \), getting \( \alpha_k > 0 \), \( x_{k+1} = x_k + \alpha_k d_k \), and \( g_{k+1} = \nabla f(x_{k+1}) \).
If \( g_{k+1} = 0 \), stop.

Step 3. Set

\[ B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k y_k} \]

where

\[ s_k = \alpha_k d_k, \]
\[ y_k = g_{k+1} - g_k. \]

Step 4. \( k := k + 1 \); go to Step 1.
The line search in Step 2 requires the steplength $\alpha_k$ to meet certain conditions. If exact line search is used, $\alpha_k$ satisfies

\begin{equation}
\min_{\alpha > 0} f(x_k + \alpha d_k).
\end{equation}

In the implementations of the BFGS algorithm, one normally requires that the steplength $\alpha_k$ satisfies the Wolfe conditions [20]:

\begin{align*}
&f(x_k + \alpha_k d_k) - f(x_k) \leq \delta_1 \alpha_k d_k^T g_k, \\
&d_k^T \nabla f(x_k + \alpha_k d_k) \geq \delta_2 d_k^T g_k,
\end{align*}

where $\delta_1 \leq \delta_2$ are constants in $(0, 1)$. For convenience, we call the line search that satisfies the Wolfe conditions (1.6)–(1.7) the Wolfe line search.

Another famous quasi-Newton method is the DFP method, which was discovered by Davidon [3] and modified by Fletcher and Powell [6]. Broyden [2] proposed a family of quasi-Newton methods:

\begin{equation}
B_{k+1}(\theta) = B_k - \frac{B_k s_k y_k^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \theta(s_k^T B_k s_k)v_k v_k^T,
\end{equation}

where $\theta \in \mathbb{R}^1$ is a scalar and $v_k = \frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k}$. The choice $\theta = 0$ gives rise to the BFGS update, whereas $\theta = 1$ defines the DFP method.

For uniformly convex functions, Powell [12] showed that the DFP algorithm with exact line searches stops at the unique minimum or generates a sequence that converges to the minimum. Dixon [4] found that all methods in the Broyden family with exact line searches produce the same iterations for general functions. For inexact line searches, Powell [14] first proved the global convergence of the BFGS algorithm with Wolfe line searches for convex functions. His result was extended by Byrd, Nocedal, and Yuan [1] to all methods in the restricted Broyden family with $\theta \in [0, 1)$. However, the following questions have remained open for many years (for example, see Nocedal [9] and Yuan [21]): (i) does the DFP method with Wolfe line searches converge for convex functions? and (ii) does the BFGS method with Wolfe line searches converge for nonconvex functions?

In this paper, we will consider the n = 2, m = 8 example in [15] for the Polak–Ribière–Polyak (PRP) conjugate gradient method [10, 11]. The two-dimensional example shows that the PRP method may cycle around eight nonstationary points if each line search picks a local minimum that provides a reduction in the objective function. By introducing a new technique of choosing parameters, we will present a new example for the PRP method (see section 2). The example has only six cyclic points. Since, in the case that $g_k^T d_k = 0$ for all $k$, the BFGS method can produce the same iterations as the PRP method does for two-dimensional functions, it can be shown by the examples that the BFGS method with Wolfe line searches need not converge for nonconvex objective functions (see section 3). Thus a negative answer is given to question (ii). The last section contains some discussions.

2. **A counterexample with six cyclic points.** The PRP method uses the negative gradient as its initial search direction. For $k \geq 1$, the method defines $d_{k+1}$ as follows:

\begin{equation}
d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k}{\|g_k\|^2} d_k.
\end{equation}
Powell [15] constructed a two-dimensional example, showing that the PRP method with the line search (2.2) may cycle around eight nonstationary points:

\[(2.2)\quad \alpha_k \text{ is a local minimum of } \Phi_k(\alpha) \text{ and such that } \Phi_k(\alpha_k) < \Phi_k(0),\]

where \(\Phi_k(\alpha)\) is the line search function

\[(2.3)\quad \Phi_k(\alpha) = f(x_k + \alpha d_k), \quad \text{where } \alpha > 0.\]

However, examples with fewer cyclic points do not seem possible from the practice in [15]. In this section, we will introduce a new technique of choosing parameters and provide an example with only six cyclic points.

Assume that \(n = 2\). Similar to [15], our example will be constructed so that all the iterations generated by the PRP method converge to the horizontal axis in \(\mathbb{R}^2\). For \(m\) even, we consider the steps \(\{s_k\}\) in the form

\[(2.4)\quad s_{mj+i} = a_i \begin{pmatrix} 1 \\ b_i \phi^{2j} \end{pmatrix}, \quad s_{mj+\overline{j}+i} = a_i \begin{pmatrix} -1 \\ b_i \phi^{2j+1} \end{pmatrix}, \quad i = 1, \ldots, \frac{m}{2},\]

where \(\phi, \{a_i\}, \{b_i\}\) are parameters to be determined, satisfying \(\phi \in (0, 1)\) and \(a_i > 0 (i = 1, \ldots, \frac{m}{2})\). To be such that

\[(2.5)\quad g_{k+1}^T d_k = 0 \quad \text{for all } k,\]

we assume that the gradients \(\{g_k\}\) have the form

\[
\begin{cases}
g_{mj+1} = c_1 \begin{pmatrix} b_{mj} \phi^{2j-1} \\ 1 \end{pmatrix}; \\ g_{mj+i} = c_i \begin{pmatrix} -b_{i-1} \phi^{2j} \\ 1 \end{pmatrix}, \quad i = 2, \ldots, \frac{m}{2}; \\ g_{mj+\overline{j}+1} = c_1 \begin{pmatrix} -b_{mj} \phi^{2j} \\ 1 \end{pmatrix}; \\ g_{mj+\overline{j}+i} = c_i \begin{pmatrix} b_{i-1} \phi^{2j+1} \\ 1 \end{pmatrix}, \quad i = 2, \ldots, \frac{m}{2},
\end{cases}
\]

where \(\{c_i\}\) are also parameters to be determined. In this section, we are interested in the case that \(m = 6\).

By relations (2.1) and (2.5), we know that the PRP method satisfies the conjugacy condition

\[(2.7)\quad d_{k+1}^T y_k = 0\]

and the descent condition

\[(2.8)\quad d_{k+1}^T g_{k+1} < 0.\]

The above conditions require that \(g_{6j+i}^T s_{6j+i} = g_{6j+i-1}^T s_{6j+i} < 0\), yielding

\[(2.9)\quad \begin{cases}
c_2(b_2 - b_1) = c_1(b_2 + b_3 \phi^{-1}) < 0, \\ c_3(b_3 - b_2) = c_2(b_3 - b_1) < 0, \\ c_1(b_1 \phi + b_3) = c_3(b_1 \phi + b_2) < 0.
\end{cases}\]

Denoting \(b_0 = -b_3 \phi^{-1}\) and \(b_4 = -b_1 \phi\), we can draw the following conditions on \(\{b_i\}\) from (2.9):

\[(2.10)\quad \begin{cases}
(b_2 - b_1)(b_1 - b_2)(b_4 - b_3) = (b_2 - b_0)(b_3 - b_1)(b_4 - b_2), \\ (b_3 - b_4)(b_2 - b_0) > 0, (b_2 - b_1)(b_3 - b_1) > 0, (b_3 - b_2)(b_2 - b_4) > 0.
\end{cases}\]
Defining $\varphi_i = b_i - b_{i-1}$, the above relations are equivalent to

\begin{equation}
\begin{cases}
\varphi_2 \varphi_3 \varphi_4 = (\varphi_1 + \varphi_2)(\varphi_2 + \varphi_3)(\varphi_3 + \varphi_4), \\
\varphi_4 (\varphi_1 + \varphi_2) < 0, \quad \varphi_2 (\varphi_2 + \varphi_3) > 0, \quad \varphi_3 (\varphi_3 + \varphi_4) < 0.
\end{cases}
\end{equation}

Further, letting $t_i = \varphi_{i+1}/\varphi_i$ and noting that $\varphi_4/\varphi_1 = -\phi$, we can obtain

\begin{equation}
\begin{cases}
t_1 t_2 t_3 = (1 + t_1) (1 + t_2) (1 + t_3) = -\phi, \\
t_1 > -1, \quad t_2 > -1, \quad t_3 < -1.
\end{cases}
\end{equation}

The first line in (2.12) is equivalent to

\begin{equation}
-t_1 t_2 t_3 = \frac{t_1 t_2 (1 + t_1) (1 + t_2)}{1 + t_1 + t_2} = \phi.
\end{equation}

Thus for any $\phi \in (0, 1)$ and $t_3 < -1$, we may solve $t_1$ and $t_2$ from (2.13). If the solved $t_1$ and $t_2$ are such that $t_1 > -1$ and $t_2 > -1$, then we can further consider the choices of $\{a_i\}$. In our real construction, we pick $t_3 = -2$. This with (2.13) indicates that

\begin{equation}
t_1 t_2 = 1 + t_1 + t_2.
\end{equation}

Further, we find that the following values of $\{t_i\}$ and $\phi$ satisfy (2.13) and allow suitable $\{a_i; i = 1, 2, 3\}$:

\begin{equation}
t_1 = \frac{3}{4}, \quad t_2 = \frac{1}{7}, \quad t_3 = -2, \quad \phi = \frac{3}{14}.
\end{equation}

Now, by the definitions of $\varphi_i$ and $t_i$, we can express $\sum_{i=2}^{4} \varphi_i$ in two ways:

\begin{equation}
\sum_{i=2}^{4} \varphi_i = b_4 - b_1 = -b_1 (1 + \phi)
\end{equation}

\begin{equation}
\varphi_2 (1 + t_2 + t_2 t_3) = (b_2 - b_1) (1 + t_2 + t_2 t_3).
\end{equation}

We then get that

\begin{equation}
b_2 = \left[1 - \frac{1 + \phi}{1 + t_2 + t_2 t_3}\right] b_1.
\end{equation}

Further, we have

\begin{equation}
b_3 = b_2 + \varphi_3 = b_2 + t_2 \varphi_2 = (1 + t_2) b_2 - t_2 b_1.
\end{equation}

Thus, letting $b_1 = 1$, we have from this, (2.17), and (2.18) that

\begin{equation}
b_1 = 1, \quad b_2 = \frac{1}{16}, \quad b_3 = \frac{5}{56}.
\end{equation}

Letting $c_2 = 1$, we obtain from (2.9) that

\begin{equation}
c_1 = -3, \quad c_2 = 1, \quad c_3 = -6.
\end{equation}

As will be shown, the parameters chosen above allow the function value to be monotonically decreased. Define $f^*$ to be the limit of $f(x_k)$. Since all the iterations
are required to converge to the horizontal axis and, for each value of the first variable, the dependence of \( f(x) \) on the second variable is linear, we have that

\[
(2.21) \quad f(x_k) - f^* = (x_k)_2 (g_k)_2 \quad \text{for all} \quad k \geq 1,
\]

where \((v)_i\) means the \(i\)th component of vector \(v\). Given the limit \( \hat{x}_1 = \lim_{j \to \infty} x_{6j+1} \), we can compute \( \{x_{6j+i}; i = 1, \ldots, 4\} \) in the following way:

\[
(2.22) \quad \begin{cases} 
  x_{6j+1} = \hat{x}_1 - \sum_{k=j}^{\infty} \sum_{i=1}^{6} s_{6k+i}, \\
  x_{6j+i} = x_{6j+i-1} + s_{6j+i-1}, \quad i = 2, 3, 4.
\end{cases}
\]

As a result, the second components of \( \{x_{6j+i}; i = 1, \ldots, 4\} \) can be expressed as follows:

\[
(2.23) \quad (x_{6j+i})_2 = -h_i(1 - \phi)^{-1}\phi^{2j}, \quad i = 1, \ldots, 4,
\]

where

\[
(2.24) \quad \begin{cases} 
  h_1 = a_1 b_1 + a_2 b_2 + a_3 b_3, \\
  h_2 = a_1 b_1 \phi + a_2 b_2 + a_3 b_3, \\
  h_3 = a_1 b_1 \phi + a_2 b_2 \phi + a_3 b_3, \\
  h_4 = h_1 \phi.
\end{cases}
\]

Using the relations (2.21) and (2.23) and noting that the structure of this example has some symmetry, we know that the monotonicity of \( f(x_k) \) requires \( \{a_i\} \) to meet

\[
(2.25) \quad -c_1 h_1 > -c_2 h_2 > -c_3 h_3 > -c_1 h_4.
\]

This relation can be satisfied if we choose

\[
(2.26) \quad a_1 = 14, \quad a_2 = 160, \quad a_3 = 1.
\]

In this case, the four terms in (2.25) have the values

\[
\frac{687}{56}, \quad \frac{387}{56}, \quad \frac{159}{28}, \quad \text{and} \quad \frac{2061}{784},
\]

respectively. So (2.25) is satisfied. Further, if we let \( (x_1)_1 = -87.5 \), then \( \{(x_{6j+i})_1; i = 1, \ldots, 6\} \) have the values \( -87.5, -73.5, 86.5, 87.5, 73.5, \) and \( -86.5 \), which are all different.

Finally, we discuss how to construct a smooth function \( f(x) \in \mathbb{R}^2 \) that satisfies the gradient conditions (2.6). At first, for given real numbers \( p_1, p_2 (\neq 0), p_3, p_4 \), and any \( j \geq 1 \), we see that the function

\[
(2.27) \quad \Psi(u_1, u_2) = \left[p_1 + p_2^{-1} p_3 (u_1 - p_1)\right] u_2
\]

is such that

\[
(2.28) \quad \nabla \Psi \left( \frac{p_1}{p_2 \phi^j} \right) = \left( \frac{p_3 \phi^j}{p_4} \right).
\]

Note that \( \{x_{6j+i}; i = 1, \ldots, 6\} \) are as follows:

\[
\begin{pmatrix} -87.5 \\ -\frac{2061}{784} \phi^{2j} \end{pmatrix}, \quad \begin{pmatrix} -73.5 \\ -\frac{387}{56} \phi^{2j} \end{pmatrix}, \quad \begin{pmatrix} 86.5 \\ -\frac{687}{56} \phi^{2j} \end{pmatrix}, \quad \begin{pmatrix} 87.5 \\ -\frac{387}{56} \phi^{2j+1} \end{pmatrix}, \quad \begin{pmatrix} 73.5 \\ -\frac{687}{56} \phi^{2j+1} \end{pmatrix}, \quad \begin{pmatrix} -86.5 \\ -\frac{2061}{784} \phi^{2j+1} \end{pmatrix}.
\]
Letting $B_i = \{u_1; |u_1 - (x_{6j+i})_1| \leq 0.1\}$, it is easy to find one-dimensional $C^\infty$ functions $\xi$ and $\gamma$ such that their values at the intervals $\{B_i; i = 1, \ldots, 6\}$ are

\[
\begin{align*}
\frac{8251}{458}, & \quad \frac{2847}{387}, \quad \frac{6981}{212}, \quad \frac{8251}{458}, \quad \frac{2847}{387}, \quad \frac{6981}{212}
\end{align*}
\]

and

\[
\begin{align*}
\frac{55}{229}, & \quad \frac{44}{387}, \quad \frac{33}{106}, \quad \frac{55}{229}, \quad \frac{44}{387}, \quad \frac{33}{106}
\end{align*}
\]

respectively. Then we can test that the function

\[(2.29) \quad f(u_1, u_2) = [\xi(u_1) + \gamma(u_1)u_1]u_2\]

is a $C^\infty$ function in $\mathbb{R}^2$ and satisfies the gradient conditions (2.6). One deficiency of the function (2.29) is that the point $x_{6j+i+1}$ may not be a local minimum of $\Phi_{6j+i}(\alpha)$ (see (2.3) for the definition of $\Phi$). For example, $x_{6j+2}$. For this, we can further choose a one-dimensional $C^\infty$ function $\tau$ such that for $i = 1, \ldots, 6$ its value at $B_i$ is equal to $(x_{6j+i})_1$. Then the $C^\infty$ function

\[(2.30) \quad f(u_1, u_2) = [\xi(u_1) + \gamma(u_1)u_1 + M(u_1 - \tau(u_1))^2]u_2\]

with $M > 0$ sufficiently large can guarantee that each $x_{6j+i+1}$ is a local minimum of $\Phi_{6j+i}(\alpha)$. This completes the construction of our new example.

Thus by introducing the quantities $\varphi_i$ and $t_i$, we have obtained a new example. The example shows that the PRP method with the line search (2.2) may cycle around six nonstationary points. One advantage of this example over the one in [15] is that it has only six cyclic points, whereas the latter has eight.

It is easy to see that the above example applies to the BFGS method if the choice of $B_1$ is such that $B_1s_1 = -lg_1$, where $l$ is any positive number. If one changes the definition of $f$ in a small neighborhood of $x_1$ to meet the necessary initial conditions, the example is also efficient for the BFGS method with any positive definite matrix $B_1$ or the PRP method with $d_1 = -g_1$.

3. Nonconvergence of the BFGS algorithm for nonconvex functions.

Generally, the line search (2.2) need not satisfy the Wolfe conditions (1.6)–(1.7). For example, consider the function

\[(3.1) \quad f(x) = \cos x, \quad x \in \mathbb{R}^1.\]

Assume that $x_k = 0$ and $d_k = 1$. For any nonnegative integer $i$, $\alpha = (2i + 1)\pi$ is a local minimum of $\Phi_{3i}(\alpha)$. Then (1.6) is false if $i$ is large. For the line search in the example of section 2, however, we can directly test that the Wolfe conditions (1.6)–(1.7) hold (see Theorem 3.1). Thus the example in section 2 also shows that the BFGS algorithm with Wolfe line searches need not converge for nonconvex objective functions.

**Theorem 3.1.** Consider the BFGS algorithm with the Wolfe line search (1.6)–(1.7), where $\delta_1 \leq \frac{60}{\sqrt{100}}$ and $\delta_2 \in (\delta_1, 1)$. Then for any $n \geq 2$ there exists a starting point $x_1$ and a $C^\infty$ function $f$ in $\mathbb{R}^n$ such that the sequence $\{\|g_k\|_2: k = 1, 2, \ldots\}$ generated by the algorithm is bounded away from zero.

**Proof.** Consider the example in section 2. For any starting matrix $B_1$, we may slightly modify the example such that it satisfies the necessary initial conditions. By (2.21), (2.23), and (2.6), we see that

\[(3.2) \quad f(x_{6j+i}) = f^* - c_i h_i (1 - \phi)^{-1} \phi^{2j}, \quad i = 1, \ldots, 4.\]
Still denote \( b_0 = -b_3 \phi^{-1} \), \( b_4 = -b_1 \phi \) and let \( a_4 = a_1 \), \( c_4 = c_1 \). We have by (2.4) and (2.6) that
\[
(3.3) \quad g^T_{6j+i}s_{6j+i} = a_1 c_1 (b_1 - b_{i-1}) \phi^{2j}, \quad i = 1, \ldots, 4.
\]
Combining (3.2) and (3.3) and noting the symmetry of the example, we know that the first Wolfe condition (1.6) holds with any constant \( \delta_1 \) satisfying
\[
\delta_1 \leq \min \left\{ \frac{f(x_{6j+i+1}) - f(x_{6j+i})}{g_{6j+i}s_{6j+i}} : i = 1, 2, 3 \right\} = \frac{1}{1 - \phi} \min \left\{ \frac{c_i b_i - c_{i+1} b_{i+1}}{a_i c_i (b_i - b_{i-1})} : i = 1, 2, 3 \right\} = \frac{69}{7480}.
\]
In addition, relations (2.5) and (2.8) imply that the second Wolfe condition (1.7) holds for \( \delta_2 \in (\delta_1, 1) \). Thus the example in section 2 shows that the BFGS algorithm with Wolfe line searches need not converge for two-dimensional functions.

In the case when \( n \geq 3 \), we need only to consider the function
\[
(3.5) \quad \hat{f}(x) = \hat{f}(u_1, u_2, \ldots, u_n) = f(u_1, u_2),
\]
where \( f \) is the function in the example of section 2. This completes our proof. \( \square \)

The parameter \( \delta_1 \) in the above theorem is required to be no greater than \( \frac{69}{7480} \approx 0.0092 \). If we consider Powell’s example with eight cyclic points, then Theorem 3.1 can be extended to \( \delta_1 \leq \frac{1}{34} = 0.02857 \).

4. Some discussions. In this paper, it has been shown by one of Powell’s examples in [15] and a new example with six cyclic points that the BFGS algorithm with Wolfe line searches need not converge for nonconvex objective functions. This result also applies to the Hestenes–Stiefel conjugate gradient method [8], the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm with Wolfe line searches, and let
\[
(4.1) \quad \begin{cases} c_2 (b_2 - b_1) = c_1 (b_2 + b_2 \phi^{-1}) < 0, \\ c_1 (b_2 + b_1 \phi) = c_2 (b_1 + b_1 \phi) < 0, \end{cases}
\]
where \( \phi \in (0, 1) \). Denote \( b_0 = -b_2 \phi^{-1} \), \( b_3 = -b_1 \phi \), \( \varphi_i = b_i - b_{i-1} \) (\( i = 1, 2, 3 \)), and \( t_i = \varphi_{i+1}/\varphi_i (i = 1, 2) \). Similar to (2.10), (2.11), and (2.12), we can obtain
\[
(4.2) \quad \begin{cases} t_1 t_2 = (1 + t_1)(1 + t_2) = -\phi, \\ t_1 > -1, \ t_2 < -1. \end{cases}
\]
The above imply that $t_2 = -(1 + t_1)$ and $\phi = t_1(1 + t_1)$. Since $\phi \in (0, 1)$, we can then get that $t_1 > 0$. Further, letting $b_1 = 1$, we can, similarly to (2.16), obtain that $b_2 = (1 + t_1)^2 / t_1$. Since $b_1, b_2$, and $\phi$ are all positive, we know by $c_1(b_2 + b_1 \phi) < 0$ that $c_1 < 0$. Letting $c_1 = -t_1$, we can get by (4.1) that $c_2 = -(1 + t_1)$. In a way similar to (2.21)–(2.25), it is easy to see that the condition $f(x_{i+1}) > f(x_{i+2})$ requires

$$c_1(a_1b_1 + a_2b_2) > -c_2(a_1b_1 \phi + a_2b_2).$$

Substituting the expressions of $\phi, c_1$, and $c_2$ with $t_1$, (4.3) is equivalent to

$$-(2 + t_1)t_1^2a_1b_1 - a_2b_2 > 0.$$

This is not possible since $t_1, a_1, a_2, b_1$, and $b_2$ are all positive. The contradiction shows the nonexistence of examples of four cyclic points.

Under the assumption that $x_k \to \bar{x}$, Powell [13] showed that the BFGS algorithm with exact line searches converges globally for general functions when there are only two variables. This result was extended by Pu and Yu [18] to the case in which $n \geq 2$. Therefore an interesting question may be, If $x_k \to \bar{x}$, is the BFGS algorithm with Wolfe line searches globally convergent for general functions? Another question is, Does there exist an inexact line search that ensures the global convergence of the BFGS method for general functions?

Recently, Powell [16] showed that if the line search always finds the first local minimum of $\Phi_k(\alpha)$ in (2.3), the BFGS method is globally convergent for two-dimensional twice-continuously differentiable functions with bounded level sets. Powell [17] and the author are trying to construct a three-dimensional example showing that the BFGS algorithm with the above line search need not converge.

Acknowledgments. The author is much indebted to Professors Y. Yuan, J. C. Gilbert, and M. J. D. Powell, who discussed with him the idea of this paper and gave many valuable suggestions and comments. Thanks are also due to the two anonymous referees, whose comments and suggestions greatly improved this paper.

REFERENCES